

Docket No.: HI-0073

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of :
Jae Shin YU, Sung Bae JUN and Kyoung Ro YOON :
Serial No.: New U.S. Patent Application :
Filed: March 7, 2002 :
For: METHOD FOR SUMMARIZING NEWS VIDEO STREAM USING
SYNTHETIC KEY FRAME BASED UPON VIDEO TEXT

#3
J1040 U.S. PRO
10/091490
03/07/02

TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Assistant Commissioner of Patents
Washington, D. C. 20231

Sir:

At the time the above application was filed, priority was claimed based on the following application:

Korean Patent Application No. 27912/2001, filed May 22, 2001.

A copy of each priority application listed above is enclosed.

Respectfully submitted,
FLESHNER & KIM, LLP



Daniel Y.J. Kim
Registration No. 36,186
Carl R. Wesolowski
Registration No. 40,372

P. O. Box 221200
Chantilly, Virginia 20153-1200
703 502-9440

Date: March 7, 2002

DYK/CRW/cng

J1040 U.S. PRO
10/091490



03/07/02

대한민국 특허청

KOREAN INTELLECTUAL PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원번호 : 특허출원 2001년 제 27912 호
Application Number PATENT-2001-0027912

출원년월일 : 2001년 05월 22일
Date of Application MAY 22, 2001

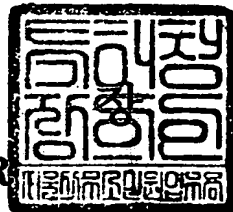
출원인 : 엘지전자주식회사
Applicant(s) LG ELECTRONICS INC.



2001 년 12 월 12 일

특 허 청

COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2001.05.22
【국제특허분류】	H04M
【발명의 명칭】	비디오 텍스트 합성 키프레임 생성방법
【발명의 영문명칭】	METHOD FOR GENERATING VIDEO TEXT SYNTHETIC KEY FRAME
【출원인】	
【명칭】	엘지전자 주식회사
【출원인코드】	1-1998-000275-8
【대리인】	
【성명】	허용록
【대리인코드】	9-1998-000616-9
【포괄위임등록번호】	1999-043458-0
【발명자】	
【성명의 국문표기】	유재신
【성명의 영문표기】	YU, Jae Shin
【주민등록번호】	770313-2465126
【우편번호】	151-014
【주소】	서울특별시 관악구 신림4동 484-31
【국적】	KR
【발명자】	
【성명의 국문표기】	전성배
【성명의 영문표기】	JUN, Sung Bae
【주민등록번호】	711010-1057913
【우편번호】	153-034
【주소】	서울특별시 금천구 시흥4동 804
【국적】	KR
【발명자】	
【성명의 국문표기】	윤경로
【성명의 영문표기】	Y00N, Kyoung Ro

【주민등록번호】 641204-1037516
【우편번호】 135-271
【주소】 서울특별시 강남구 도곡동 도곡 경남아파트 101동 2004호
【국적】 KR
【심사청구】 청구
【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사 를 청구합니다. 대리인 허용록 (인)
【수수료】
【기본출원료】 20 면 29,000 원
【가산출원료】 7 면 7,000 원
【우선권주장료】 0 건 0 원
【심사청구료】 16 항 621,000 원
【합계】 657,000 원
【첨부서류】 1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 내용기반 멀티미디어 검색 및 브라우징에 관한 것으로서, 특히 합성된 비디오 텍스트 키프레임을 이용해서 뉴스 비디오 콘텐츠를 요약할 수 있도록 한 비디오 텍스트 합성 키프레임 생성방법에 관한 것이다.

본 발명은 뉴스 비디오 콘텐츠에 대한 텍스트 키프레임 기반의 요약과 비선형적인 브라우징을 위하여, 뉴스 비디오 콘텐츠가 샷 단위로 분할되고 앵커 샷과 에피소드 샷으로 구분된 뉴스 비디오에서 텍스트가 검출된 콘텐츠에 대해; 검출된 텍스트가 앵커 샷에서 추출된 텍스트인지 혹은 에피소드 샷에서 추출된 텍스트인지에 따라 텍스트 중요도 가중치를 다르게 할당하고, 에피소드 샷에서 추출된 텍스트 영역들에 대하여는 텍스트 영역의 크기와 지속시간, 평균 텍스트의 크기 등을 고려하여 각각의 텍스트 영역의 중요도 가중치를 할당하며, 검출된 텍스트 중에서 일정 가중치 이상의 텍스트를 선택하고 이 것을 하나의 이미지에 합성하여 하나의 뉴스 기사 구간을 대표하는 텍스트 합성 키프레임으로 생성함으로써, 이 텍스트 합성 키프레임을 기반으로 뉴스 비디오 콘텐츠를 요약할 수 있도록 하였다.

【대표도】

도 6

【색인어】

멀티미디어, 비디오, 뉴스, 콘텐츠, 요약, 텍스트, 합성 키프레임

【명세서】

【발명의 명칭】

비디오 텍스트 합성 키프레임 생성방법{METHOD FOR GENERATING VIDEO TEXT
SYNTHETIC KEY FRAME}

【도면의 간단한 설명】

도1은 비디오 스트림의 구조적 정보를 설명하기 위한 도면

도2는 합성 키프레임의 개념을 설명하기 위한 도면

도3은 텍스트 합성 키프레임의 개념을 설명하기 위한 도면

도4는 뉴스 합성 키프레임을 이용한 비디오 브라우징 인터페이스의 예를 나타낸 도면

도5는 비디오에서 텍스트를 추출하는 방법을 설명하기 위한 도면

도6은 뉴스 비디오에서 텍스트 합성 키프레임의 개념을 설명하기 위한 도면

도7은 추출된 텍스트 영역 내의 글자 크기를 예측하는 방법을 나타낸 도면

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

<8> 본 발명은 내용 기반 멀티미디어 검색과 브라우징에 관한 것으로서 특히, 비디오 콘텐츠(video content)의 검색(searching/filtering)과 브라우징(browsing)을 위하여 비디오 콘텐츠에서 추출되는 텍스트(text)를 이용해서 텍스트 합성 키프레임(keyframe)을 생성하고, 이 텍스트 합성 키프레임을 이용해서

뉴스 비디오를 요약할 수 있도록 한 비디오 텍스트 합성 키프레임 생성방법에 관한 것이다.

- <9> 디지털 비디오 기술의 발달과 이미지/비디오/오디오 인식 기술의 발달로 인하여 사용자들은 원하는 멀티미디어 콘텐츠를 원하는 시점에 원하는 부분만을 검색(searching/filtering)하고 브라우징(browsing) 할 수 있게 되었다.
- <10> 비선형적인 멀티미디어 콘텐츠 브라우징(non-linear multimedia content browsing)과 검색을 위하여 가장 기본이 되는 기술은 샷 분할(shot segmentation) 기법과 샷 클러스터링(shot clustering) 기법이며, 이 두 가지 기술은 멀티미디어 콘텐츠를 구조적으로 분석하는데 있어서 가장 핵심이 되는 기술이다.
- <11> 도1을 참조하여 비디오 스트림(video stream)의 구조적 정보의 이해를 돕는다. 비디오에서 샷(shot)이란 방해(interruption)없이 하나의 카메라로부터 얻어진 비디오 프레임들의 시퀀스(sequence)를 의미하며, 이는 비디오를 분석(analysis)하거나 구성(construction)하는 가장 기본이 되는 단위이다.
- <12> 또한 일반적으로 비디오에는 의미적인 구성 요소인 씬(scene)이 존재하는데, 여기서 씬은 이야기의 전개 또는 비디오 구성에 있어서 의미적인 구분 요소이며, 일반적으로 하나의 씬 내에는 여러 개의 샷이 존재한다. 이러한 샷과 씬의 개념은 비디오뿐만 아니라 오디오 프로그램에도 같은 방식으로 적용 가능하다.
- <13> 궁극적으로 멀티미디어 콘텐츠 인덱싱 기법은 콘텐츠에 존재하는 샷/씬 단위의 구조적인 정보를 추출하고 각각의 구조 단위 별로 그 세그먼트를 대표할 수

있는 키 프레임 등과 같은 주요 특징 요소를 추출하여 멀티미디어 콘텐츠에 대한 구조적 정보를 인덱싱하고, 사건(event)의 발생, 시청각 오브젝트(audio-visual object)의 등장, 오브젝트의 상태, 배경(장소 등) 등에 대한 의미적인 정보를 시간축 위에 기술하는 방식으로 인덱싱하여 사용자가 원하는 부분을 쉽게 브라우징하거나 검색할 수 있도록 하는데 그 목적이 있다.

<14> 다른 한편으로 상기한 바와 같이 효율적인 비디오 검색과 브라우징을 위하여 비디오 스트림에서 의미적 정보를 표현하는 주요 텍스트 영역이나 뉴스 아이콘 또는 사람의 얼굴 영역 등을 추출하기 위한 연구도 진행되고 있으며, 이러한 키 영역을 합성하여 새로운 키프레임을 추출하는 방법도 소개되고 있다. 합성 키프레임의 경우는 썸이나 샷 등의 세그먼트를 대표할 수 있는 키 영역이나 키프레임을 추출하고 추출된 키 영역이나 키프레임을 이용해서 이들을 하나의 이미지로 합성함으로써, 이 합성된 이미지가 해당 세그먼트의 내용을 대표하는 수단으로 제공되는 방법이다. 이렇게 합성 키프레임을 이용함으로써 작은 디스플레이 공간에 많은 정보를 함축적으로 표현할 수 있고, 사용자는 콘텐츠의 특정 부분의 내용을 보다 빠르고 쉽게 이해할 수 있을 뿐만 아니라, 관심이 있는 부분만 선택하여 디스플레이(시청)하는 것도 가능하게 되었다.

<15> 멀티미디어 콘텐츠에서 일반적으로 뉴스 비디오 스트림은 정형화/구조화된 비디오 데이터로서 공간적/시간적 구조를 가지는 정형화된 모델을 따른다.

<16> 즉, 뉴스 비디오 스트림은 일반적인 멀티미디어 스트림과는 달리 정형화/구조화된 비디오 데이터로서, 하나의 뉴스 비디오 스트림은 여러 개의 기사

(article)로 구성되어 있고, 하나의 기사 내에서는 뉴스 앵커(anchor)의 기사 요약 설명 부분과 기사 내용을 뒷받침하는 에피소드(episode) 부분으로 구성된다.

<17> 즉, 하나의 뉴스 비디오 스트림은 여러 개의 기사(article)로 구성되어 있고, 하나의 기사 내에서는 뉴스 앵커의 기사 요약 설명 부분(앵커 샷)과 기사 내용을 뒷받침하는 자료화면(에피소드) 부분으로 구성된 것을 보여준다. 내용상으로 볼 때 종합 뉴스의 경우 정치, 경제, 사회, 스포츠, 날씨에 관한 사항들이 총망라되며, 다른 장르의 비디오 콘텐츠와는 달리 뉴스 비디오는 정형적인 구조를 가지며, 사용자에게 따라 관심이 있는 기사의 차이가 분명하므로 사용자는 실제로 원하는 뉴스 기사만을 빠르게 검색하고자 하는 요구가 많다.

<18> 상기한 바와 같이 사용자가 실제로 원하는 뉴스 기사만을 빠르게 검색하고자 하는 요구를 비디오 인덱싱 관점에서 수용하기 위하여 뉴스 비디오의 구조적/의미적 정보를 사용하여 뉴스 비디오 스트림을 기사 단위로 인덱싱하기 위한 연구가 지속되고 있다.

<19> 즉, 뉴스 비디오의 시간적 구조를 이용하여 기사 단위로 인덱싱된 뉴스 비디오 데이터에 TOC(Table Of Content)나 스토리보드(storyboard)와 같은 인터페이스를 결합한 여러 가지 비선형적 뉴스 비디오 브라우징 기법들이 연구되고 있으나 뉴스비디오에 대하여 각각의 기사(씬)를 대표할 만한 키 프레임을 선정하는 것이 매우 어렵고, 단순 스토리보드 형태의 요약방법은 사용자에게 실제 이야기 단위인 씬에 대한 정보를 효율적으로 전달할 수 없기 때문에 뉴스 기사의 요약 방식에 있어서 매우 비효율적이며, 사용자에게 전체 뉴스 내용을 직관적으로 전달하기 어려운 단점이 있다.

【발명이 이루고자 하는 기술적 과제】

- <20> 본 발명은 뉴스 비디오를 검색하거나 브라우징함에 있어서, 뉴스 기사를 기반으로 하여 보다 효율적으로 뉴스 비디오 내용을 요약하고, 이 요약된 내용을 브라우징 인터페이스로 하여 사용자가 희망하는 뉴스 기사를 디스플레이 해줄 수 있도록 하기 위하여 텍스트 합성 키프레임을 생성하는 방법을 제공한다.
- <21> 본 발명은 뉴스 비디오 콘텐츠의 의미적/구조적 정보를 활용하여 의미 전달성이 높은 뉴스 비디오에서의 합성 텍스트 키프레임을 생성하고, 이 합성 텍스트 키프레임을 이용해서 기사 기반의 뉴스 비디오 요약을 보다 효율적으로 수행할 수 있도록 한 비디오 텍스트 합성 키프레임 생성방법을 제안한다.
- <22> 본 발명은 뉴스 비디오에 대한 합성 키프레임을 생성함에 있어서, 뉴스 비디오 콘텐츠에서 검출된 텍스트가 앵커 샷에 속한 것인지 혹은 에피소드 샷에 속한 것인지에 따라 다른 중요도를 할당하여 비디오 텍스트 합성 키프레임을 생성함으로써, 사용자에게 더욱더 직관적인 뉴스 비디오 요약 환경을 제공할 수 있도록 한 비디오 텍스트 합성 키프레임 생성방법을 제안한다.
- <23> 본 발명은 뉴스 비디오에 대한 합성 키프레임을 생성하고 이 합성 키프레임을 이용해서 뉴스 비디오 콘텐츠를 요약함에 있어서, 뉴스 비디오 콘텐츠에서 검출된 텍스트가 앵커 샷에 속한 것인지 혹은 에피소드 샷에 속한 것인지, 에피소드 샷에 속한 텍스트인 경우에는 그 텍스트 영역의 크기와 지속시간, 텍스트 영역에서의 평균 텍스트 요소의 크기가 어느 정도인지 등에 따라 서로 다른 중요도가중치를 할당하는 방법으로 텍스트 기반 합성 키프레임을 생성함으로써, 사용

자에게 더욱더 직관적인 뉴스 비디오 요약 환경을 제공할 수 있도록 한 비디오 텍스트 합성 키프레임 생성방법을 제안한다.

【발명의 구성 및 작용】

- <24> 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 뉴스 장르의 구조적/의미적 특성을 이용해서 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.
- <25> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 뉴스 콘텐츠에서 구조적/의미적으로 중요한 의미를 갖는 앵커 샷 텍스트를 이용해서 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.
- <26> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 뉴스 콘텐츠에서 특정 구간에 존재하는 텍스트의 일반적인 중요도와 함께 뉴스의 구조적/의미적 특징에 기인하는 앵커 샷 텍스트의 중요도를 함께 고려하여 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.
- <27> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 뉴스 콘텐츠에서 구조적/의미적으로 중요한 의미를 갖는 앵커 샷과 에피소드 샷의 텍스트를 이용해서 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.
- <28> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 상기 앵커 샷과 에피소드 샷의 텍스트에 각각 다른 중요도 가중치를 할당하여 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.
- <29> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 상기 앵커 샷과 에피소드 샷의 텍스트에 각각 다른 중요도 가중치를 할당하고, 상기 에피소드 샷

텍스트에 대해서는 에피소드 샷 텍스트 영역의 특징요소들에 대한 중요도 가중치를 다르게 할당하여 에피소드 샷 텍스트의 중요도를 구함으로써, 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.

<30> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법에서 상기 에피소드 샷 텍스트 영역에 대한 중요도 가중치를 할당함에 있어, 상기 에피소드 샷 텍스트 영역의 크기, 지속시간, 텍스트 영역에서의 평균 텍스트의 크기 등을 고려하여 텍스트 합성 키프레임을 생성하는 것을 특징으로 한다.

<31> 또한 본 발명의 비디오 텍스트 합성 키프레임 생성방법은 뉴스 콘텐츠의 특정 구간에 존재하는 텍스트를 검출하는 단계, 상기 검출된 텍스트가 앵커 샷 텍스트인지 혹은 에피소드 샷 텍스트인지를 판단하는 단계, 상기 판단 결과에 따라 텍스트 중요도 가중치를 차별하여 할당하는 단계, 상기 검출된 텍스트 중에서 상기 가중치에 따라 선택된 텍스트를 하나의 이미지에 합성하여 텍스트 합성 키프레임을 구성하는 단계로 이루어지는 것을 특징으로 한다.

<32> 첨부된 도면을 참조하여 상기 본 발명의 비디오 텍스트 합성 키프레임 생성방법을 더욱 상세하게 설명한다.

<33> 앞서 설명한 바와 같이 효율적인 비디오 검색과 브라우징을 위하여, 씬이나 샷 등의 세그먼트를 대표할 수 있는 키 영역이나 키프레임 영역을 추출해서 하나의 합성된 이미지를 생성하고 이 합성된 이미지를 합성 키프레임으로 하여 해당 세그먼트의 내용을 대표하도록 하는 방법이 있다.

<34> 도2는 이와 같은 합성 키프레임의 개념을 보여준다. 이 방법은 추출된 비디오 스트림의 구조적 구성요소인 씬과 샷을 대표하기 위하여 키프레임을 추출하고 이를 검색이나 브라우징을 위한 용도로 사용하는 방법으로서, 비디오의 논리적 단위인 씬 혹은 물리적 단위인 샷에서 키프레임을 추출하고 추출된 키프레임을 이용해서 논리적 또는 물리적으로 합성하여 구성한 합성 키프레임을 사용자에게 제공함으로써 쉽게 비디오 콘텐츠를 이해하고 또 원하는 위치로 빠르게 접근할 수 있도록 하고 있다.

<35> 합성 키프레임을 생성함에 있어 텍스트를 중심으로 합성 키프레임을 생성하고, 텍스트 합성 키프레임을 이용해서 비디오 검색과 브라우징을 수행할 수도 있다. 도3은 텍스트 합성 키프레임의 개념을 보여준다. 도3에 나타낸 바와 같이, 비디오 콘텐츠의 내용을 함축적으로 전달하는 요소로 많이 사용되는 텍스트 영역을 추출하고, 추출된 텍스트의 지속시간, 크기 등의 정보를 이용해서 중요도를 판별하고, 중요도가 높은 텍스트들을 합성하여 합성 키프레임을 생성하고, 이렇게 구한 텍스트 합성 키프레임을 사용자에게 제공함으로써, 텍스트를 기반으로 하는 비디오 콘텐츠의 검색과 브라우징을 가능하게 하는 것이다.

<36> 이 방법에서 텍스트의 중요도는 텍스트의 디스플레이 지속시간이나 추출된 텍스트의 크기뿐만 아니라, 추출된 텍스트 영역 내의 텍스트 요소의 평균 크기 등을 고려할 수 있으며, 이러한 중요도를 고려해서 선택한 몇 개의 텍스트를 합성하여 합성 텍스트 키프레임을 생성하는 것이다. 이와 같이 중요도를 이용한 텍스트 합성 키프레임은 비디오 콘텐츠의 내용을 중요도가 높은 텍스트를 이용해서

요약함으로써 사용자의 이해를 돕고, 비디오의 내용을 함축하여 보여줄 수 있는 기능을 제공한다.

<37> 본 발명은 상기 합성 텍스트 키프레임을 생성함에 있어서 합성 텍스트 키프레임이 얼마나 정확하게 그 비디오 구간을 함축하여 의미적으로 충분한 요약과 전달을 할 수 있도록 생성할 수 있는가의 문제를 해결하며, 특히 뉴스 비디오에서 뉴스 비디오 콘텐츠의 효율적인 검색과 브라우징을 위한 합성 텍스트 키프레임을 생성하는 방법을 제시한다.

<38> 앞서 설명한 바와 같이, 일반적으로 뉴스 비디오 스트림은 정형화/구조화된 비디오 데이터로서 공간적/시간적 구조를 가지는 정형화된 모델을 따른다. 하나의 뉴스 비디오 스트림은 여러 개의 기사(Article)로 구성되어 있고, 하나의 기사 내에서는 뉴스 앵커의 기사 요약설명 부분(앵커 샷)과 기사 내용을 뒷받침하는 자료화면(에피소드) 부분으로 구성된다. 내용상으로 볼 때 종합 뉴스의 경우 정치, 경제, 사회, 스포츠, 날씨에 관한 사항들이 총 망라되며 다른 장르의 비디오 콘텐츠와는 달리 뉴스비디오는 정형적인 구조를 가진다.

<39> 대부분의 비디오 인덱싱 시스템에서는 추출된 비디오 스트림의 구조적 구성 요소인 씬과 샷을 대표하기 위하여 키 프레임(key frame)을 추출하고, 이를 검색이나 브라우징을 위한 용도로 사용한다. 이때 키 프레임은 단위 세그먼트를 잘 표현하기 위하여 비디오 시퀀스에서 추출된 이미지 프레임을 뜻한다(도2 및 도3 참조).

<40> 키 프레임으로 비디오 스트림의 샷이나 씬을 대표하는 방식 이외에 뉴스 비디오의 인덱싱과 같은 특정 어플리케이션 분야에 적용 가능한 수단으로 키 영역

을 이용한 비디오 브라우징 인터페이스가 제안되었다. 이 기술에서는 앵커 샷(anchor shot)에 존재하는 뉴스 아이콘(news icon)을 키영역(key region)으로 선정하여 이를 추출하고 해당 뉴스 아이템을 대표하는 수단으로 이용하였다. 일반적으로 키 영역은 텍스트(text), 사람 얼굴, 뉴스 아이콘 등 비디오 세그먼트의 내용을 함축하는 영역을 가리키며, 이러한 키 영역은 비디오 스트림의 의미적 정보를 나타내므로 비디오 요약이나 비디오 스트림에 대한 비 선형적 브라우징 인터페이스를 위하여 사용된다. 비디오에서 텍스트를 추출하는 방법이나 사람의 얼굴영역을 추출하는 방법, 뉴스 아이콘을 추출하는 방법은 이미지에서 특정 부분의 칼라 분포 특성이나 가로/세로 비율, 움직임(모션:motion) 정보 등을 종합적으로 사용한다.

<41> 도4는 특히, 합성 키프레임을 이용하는 뉴스 비디오 브라우징 인터페이스의 예를 잘 보여주고 있다.

<42> 뉴스 진행자인 앵커가 등장하는 장면과 뉴스 아이콘, 그리고 앞서 설명한 텍스트 합성 키프레임, 대표 장면(일반적 키프레임)을 보여준다.

<43> 도5는 비디오 텍스트를 추출하는 방법을 개념적으로 보여준다. 비디오에 나오는 텍스트의 최소 사각형(MBR) 영역만 추출하여 사용자에게 제시해 줌으로써, 비디오의 내용을 쉽게 파악할 수 있도록 해주고, 또한 인덱스할 수 있는 기능을 제공하며, 유연성 있는 정보 검색과 인덱싱된 정보를 이용해서 네트워크 상에서 원거리 정보 검색도 가능하게 한다.

<44> 이와 같이 비디오에서 텍스트 영역을 추출하는 방법은, 가로와 세로 엣지 히스토그램(edge histogram)이 집중적으로 나타나는 성질과, 엣지 히스토그램이

글자 간격에 따라 반복적으로 커졌다 작아졌다 하는 정보를 이용해서 텍스트 후보 영역을 추출하고, 텍스트 후보 영역 중에서 가로 세로 비율이 텍스트임을 만족할 만한, 모션이 적고 배경과 명도차가 큰 칼라가 사용된 영역을 텍스트로 추출하는 방법을 이용한다.

<45> 기존의 텍스트 합성 키프레임 추출방법은 비디오 스트림 내에 존재하는 텍스트에 대하여 텍스트 영역의 크기, 추출된 영역 내의 텍스트의 평균 크기, 디스플레이 지속시간을 이용해서 중요도를 계산하고, 중요도가 높은 몇 개의 텍스트를 선택하여 하나의 키프레임으로 합성함으로써 텍스트 합성 키프레임을 생성하고, 브라우저에서는 텍스트 합성 키프레임을 이용해서 해당 구간을 대표하는 수단으로 이용하였다.

<46> 그런데, 이러한 단순한 중요도 판별은 장르의 특성이나 비디오 콘텐츠의 구조적/의미적 정보를 활용하지 않기 때문에 실제로는 중요한 의미를 전달하는 텍스트 요소임에도 불구하고 합성 키프레임 생성에서 제외될 가능성이 매우 높다. 따라서, 각 장르의 구조적/의미적 특성을 파악하여 그 장르에서는 어떤 텍스트가 중요한 텍스트인지 판단하는 것은 중요도를 이용한 텍스트 합성 키프레임 생성에 있어서 매우 중요한 문제이다. 본 발명에서는 뉴스 비디오 콘텐츠의 구조적/의미적 정보를 활용해서 의미 전달성이 높은 텍스트 합성 키프레임 생성방법을 제안하고자 하는 것이다.

<47> 앞서 설명한 바와 같이 뉴스 비디오에서는 전체적인 구조가 앵커 샷과 에피소드 샷의 한 기사(article)가 여러 개 있는 구조로 정형화되어 있다. 뉴스에서 앵커 샷은 뒤에 이어지는 에피소드 샷을 요약해 주는 역할을 하는 경우가 거의

대부분이다. 예를 들면 앵커 샷 내에 나타나는 텍스트는 하나의 뉴스 아이템을 요약하여 함축적으로 표현(기술)하는 경우가 대부분이다. 따라서, 텍스트 키 프레임만으로 뉴스를 요약하고자 할 때 앵커 샷에 나오는 텍스트는 한 기사를 요약할 수 있기 때문에 매우 중요한 텍스트 정보가 될 것이다.

<48> 그러므로, 본 발명에서는 앵커 샷 텍스트에 대해서는 에피소드 샷에 나오는 텍스트에 비해서 더 높은 중요도 가중치를 주는 방법으로 텍스트 합성 키프레임을 생성한다.

<49> 이와 같은 방법으로 뉴스 비디오의 구조적/의미적 정보를 고려해서 텍스트 합성 키프레임을 생성함으로써, 적은 양의 데이터로 많은 정보를 제공하여 뉴스를 요약하고, 비선형적인 뉴스 비디오의 수단으로 이용할 수 있게 하였다.

<50> 도6은 본 발명에 따른 뉴스 비디오 콘텐츠에서의 텍스트 합성 키프레임의 개념과 추출방법을 보여준다.

<51> 뉴스 텍스트 합성 키프레임은 뉴스의 한 기사(Article) 단위로 존재하는 앵커 샷(Anchor shot) 텍스트와 일반 텍스트의 키 영역을 추출하고, 추출된 영역에 대해서 앵커 샷 텍스트는 더 높은 가중치를 주고, 에피소드 샷(Episode shot)에서 추출된 일반 텍스트는 중요도를 계산하여 가중치를 준다. 이렇게 해서 아이콘 텍스트와 일정 가중치 이상의 일반 텍스트를 한 기사 단위로 합성하여 새로운 텍스트 합성 키프레임을 생성한다.

<52> 여기서 중요도 판단 함수로 고려할 수 있는 것으로는, 아이콘 텍스트(ICON TEXT)인가, 추출된 텍스트 영역의 크기는 어느 정도인가, 또 텍스트가 얼마나 지

속적으로 디스플레이되고 있는가, 추출된 텍스트 영역내의 평균 텍스트 크기가 얼마나 큰가 작은가 등이 될 수 있다. 이러한 중요도 함수를 적용해서 한 기사(Article) 단위로 아이콘 텍스트(Icon Text)와 일반 텍스트(General Text)를 적절하게 선택하여 합성 키프레임을 생성하며, 이 때 일반 텍스트는 중요도가 높은 것을 몇 개 선택하여 합성 텍스트 키프레임에 넣는다.

<53> 이 방법을 좀 더 상세하게 살펴본다.

<54> 뉴스 장르는 앵커 샷과 에피소드 샷의 기사 단위로 구조화되어 있으므로, 앵커 샷과 에피소드 샷의 구분은 앵커 샷의 검출로 가능하다. 앵커 샷의 검출은 얼굴 검출(face detection), 모션(motion) 정보, 코딩되지 않은 블록(Not coded macro block)의 비율, 칼라정보 등을 이용해서 자동으로 검출할 수 있으며, 이것은 반자동 또는 수동 작업에 의해서도 이루어질 수 있다. 즉, 앵커 샷이 갖는 특징적 요소들, 앵커 샷에서는 앵커의 얼굴이 클로즈업되어 등장한다던가, 모션이 작다던가 하는 등의 특징을 이용해서 앵커 샷을 검출할 수 있고, 앵커 샷이 검출되면 상대적으로 에피소드 샷 또한 검출될 것이다. 앵커 샷이 검출되면 전체 뉴스가 앵커 샷과 에피소드 샷으로 구분될 수 있고, 앵커 샷 내에 나타나는 텍스트는 하나의 뉴스 아이템을 요약하는 텍스트가 대부분이므로, 앵커 샷에서 검출된 텍스트인 경우에는 해당 텍스트에 대해서 가장 큰 중요도를 부여하고, 나머지 에피소드 샷에 존재하는 텍스트들은 다음과 같이 중요도를 계산한다.

<55>
$$I = A * a + B * b + C * c \quad (a + b + c = 1).$$

- <56> 여기서, I는 중요도, A는 추출된 텍스트 영역의 크기, B는 추출된 텍스트 영역 내의 텍스트 요소의 평균 크기, C는 텍스트의 디스플레이 지속시간(존속시간), a,b,c는 각각의 가중치이다.
- <57> 위와 같이 하여 중요도를 얻을 수 있다.
- <58> 에피소드 샷 내의 텍스트에 가중치를 주는 방법은 추출된 텍스트 영역의 최소 사각형(MBR)의 크기가 크고, 추출된 영역 내의 텍스트 요소의 평균 크기가 크고, 텍스트의 디스플레이 지속시간이 길면 가중치를 크게 준다. 반대의 경우 즉, 추출된 텍스트의 크기가 작고, 평균크기가 작고, 디스플레이 지속시간이 짧은 경우에는 가중치를 작게 준다.
- <59> 추출된 텍스트 영역 내의 텍스트 요소의 평균 크기는 도7에 도시된 바와 같이 히스토그램의 밀도와 크기로 알 수 있다. 즉, 글자 크기가 작으면 각 라인 사이에 가로 엷지 히스토그램의 크기가 크게 작아지는 것을 볼 수 있고, 세로 엷지 히스토그램도 각 글자 사이에서 크게 작아지는 것을 볼 수 있다. 이와 반대로, 추출된 영역 내에 글자 크기가 크면 가로 엷지 히스토그램의 분포가 넓게 분포되어 있고, 중간에 히스토그램의 크기가 갑자기 줄어드는 현상은 보이지 않는다. 이와 같은 히스토그램의 밀도와 크기 정보를 이용해서, 추출된 텍스트 영역 내의 텍스트 요소의 평균 크기를 구할 수 있다.
- <60> 추출된 텍스트의 지속시간은 이전에 추출된 영역과 현재 추출된 영역을 비교해서 구할 수 있다. 추출된 영역의 크기와 위치 정보가 유사하고, 엷지 히스토그램의 차이가 일정 임계치(Threshold Value) 보다 작으면 두 영역을 같은 텍스트로 판단하여, 추출된 텍스트의 디스플레이 지속시간을 늘린다.

- <61> 위와 같이 해서 생성한 텍스트 합성 키프레임은 뉴스 비디오 스트림을 구성하고 있는 논리적 단위인 한 기사(Article) 단위로 생성할 수 있다. 한 기사는 앵커 샷을 기준으로 하나의 기사로 구분한다.
- <62> 한편, 위와 같이 합성할 텍스트는 디스플레이할 디바이스의 크기나 브라우저에서 텍스트 합성 키프레임의 영역에 따라 그 개수나 각각의 사이즈를 늘이거나 줄일 수 있다. 예를 들어, 디스플레이할 영역이 커지면 합성할 텍스트의 수를 늘이거나 각각의 텍스트의 사이즈를 늘일 수 있고, 디스플레이할 영역이 작아지면 합성할 텍스트의 개수를 줄이거나 각각의 텍스트의 사이즈를 줄일 수 있다.
- <63> 이와 같이 합성할 텍스트의 개수의 변경이나 사이즈의 변경은 사용자의 가독성(readability)을 고려한다.
- <64> 지금까지 뉴스 비디오 콘텐츠의 구조적/의미적 정보를 이용해서 텍스트 합성 키프레임을 생성하는 방법을 설명하였다.
- <65> 상기 본 발명에 의해서 생성되는 텍스트 합성 키프레임의 어플리케이션(application)으로는 UMA 응용 어플리케이션(Universal Multimedia Access Applications)이 있다. 이 UMA 응용 어플리케이션에 대해서 살펴본다.
- <66> 사용자의 단말장치나 사용자 단말장치와 서버를 연결하는 네트워크 환경 등에 따라 사용자가 이용할 수 있는 데이터에 한계가 있을 수 있다. 즉, 어떤 디바이스를 사용하는가에 따라, 정지영상은 지원되지만 동영상은 지원되지 않을 수도 있으며, 오디오는 지원되지만 영상은 지원되지 않을 수도 있다. 또한 네트워크 연결방식이나 매체에 따라 전송용량이 부족하여 정해진 시간 내에 전송받을 수

있는 데이터의 양이 제한될 수도 있다. 이러한 다양한 사용자의 환경 변화에 적응하여, 어떤 멀티미디어 정보를 사용자의 환경에 가장 알맞은 형태로 가공하여 사용자의 편의를 도모하고, 정보의 전달 능력을 향상시키고자 하는 모든 어플리케이션들을 Universal Multimedia Access Applications라고 한다.

<67> 예를 들어 디바이스나 네트워크 등의 제약조건에 의해 비디오 스트림을 받아 디스플레이할 수 없는 사용자에게 사용자의 환경이 허락하는 한도 내의 축소된 크기와 개수의 텍스트 키프레임으로 비디오 스트림을 변화시켜 전송해 준다면 해당 비디오 콘텐츠에 대한 최소한의 이해를 도모할 수 있다. 본 발명의 텍스트 합성 키프레임은 UMA 어플리케이션에 적용될 때, 전송될 키프레임의 개수를 줄여서 전송 데이터의 양을 줄이면서도 더 많은 의미적 정보를 제공하는 수단으로 이용될 수 있다.

<68> 본 발명과 관련된 다른 어플리케이션으로는 이미 설명한 바와 같이 뉴스의 비선형적인 비디오 브라우징 어플리케이션이 있다. 전체 뉴스 비디오 스트림은 요약물 하지 않으면 사용자가 그 뉴스 비디오 스트림을 이해하기 위하여 해당 비디오 스트림 전체를 모두 시청해야 하는 단점이 있다. 또 원하는 위치로 이동하려고 해도 뉴스 비디오 스트림 내의 그 위치까지 사용자가 직접 탐색해야 하기 때문에 많은 시간이 필요하다. 따라서, 비디오 스트림을 빠른 시간 내에 검색하고 접근하기 위해서는 비선형적인 비디오 브라우징이 이용될 수 있는데, 본 발명에 따르면 전체 비디오의 내용을 텍스트 합성 키프레임으로 대표하여 특정 범위 단위로 요약하여 보여주고 그 원하는 위치부터 검색하는 것이 가능하게 된다.

【발명의 효과】

<69> 본 발명에서는 뉴스 비디오 콘텐츠에서 구조적/의미적으로 중요한 의미를 갖는 앵커 샷 텍스트를 이용해서 합성 키프레임을 생성함으로써 사용자가 뉴스 비디오 콘텐츠를 보다 더 명확하게 요약하여 이해할 수 있는 장점이 있다.

<70> 또한 본 발명에서는 뉴스 비디오 콘텐츠에서 구조적/의미적으로 중요한 의미를 갖는 앵커 샷 텍스트와, 에피소드 샷 텍스트를 구분하여 각각 다른 중요도를 기반으로 합성 키프레임을 생성함으로써 사용자가 뉴스 비디오 콘텐츠를 보다 더 명확하게 요약하여 이해할 수 있는 장점이 있다.

1020010027912

출력 일자: 2001/12/13

【특허청구범위】**【청구항 1】**

뉴스 비디오에서 검출된 텍스트가 앵커 샷 텍스트인지, 에피소드 샷 텍스트인지를 판별하는 단계, 상기 뉴스 비디오에서 검출된 텍스트가 앵커 샷 텍스트인 경우와 에피소드 샷 텍스트인 경우에 대해서 텍스트 중요도를 각각 차별하여 할당하는 단계, 상기 할당된 텍스트 중요도를 기준으로 하나 이상의 텍스트를 선택하고 이 선택된 텍스트를 하나의 이미지에 합성하여 텍스트 합성 키프레임을 생성하는 단계를 포함하여 이루어지는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 2】

제 1 항에 있어서, 상기 앵커 샷 텍스트인 경우를 에피소드 샷 텍스트인 경우보다 더 높은 중요도를 할당하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 3】

제 1 항에 있어서, 상기 에피소드 샷 텍스트에 대해서는 텍스트 영역의 크기와 디스플레이 지속시간 및 텍스트 영역에서의 텍스트 요소의 평균 크기를 고려하여 중요도를 할당하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 4】

제 1 항 내지 제 3 항 중의 어느 한 항에 있어서, 상기 텍스트 중요도는 중요도에 대한 가중치를 달리하여 적용함으로써, 텍스트 중요도를 차별하여 구하고 할당하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 5】

제 4 항에 있어서, 상기 에피소드 샷 텍스트의 중요도는 텍스트 영역의 크기와 지속시간, 텍스트 요소의 평균 크기에 비례하여 할당하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 6】

제 5 항에 있어서, 상기 텍스트 영역의 크기는 텍스트 영역을 감싸는 최소 사각형(MBR)의 크기로 판단하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 7】

제 5 항에 있어서, 상기 텍스트 영역에서의 텍스트 요소의 평균 크기는 추출된 텍스트 영역 내에서 엣지의 가로, 세로 히스토그램의 밀도와 크기를 이용해서 판단하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 8】

제 5 항에 있어서, 상기 텍스트 영역의 디스플레이 지속시간은 이전에 검출된 텍스트 영역과 현재 검출된 텍스트 영역의 유사성을 판단하여 같은 텍스트라

고 판단된 경우 해당 텍스트에 대한 디스플레이 지속시간을 늘리는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 9】

제 5 항에 있어서, 상기 텍스트 영역의 크기, 지속시간, 텍스트 영역에서의 텍스트 요소 평균 크기에 비례하여 가중치를 높이는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 10】

제 1 항에 있어서, 상기 텍스트 합성 키프레임 생성을 위하여 선택되는 텍스트의 개수와 크기는 텍스트 합성 키프레임을 디스플레이하는 디바이스나 브라우저의 크기를 고려하여 결정하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 11】

제 10 항에 있어서, 상기 디바이스나 브라우저의 크기가 크고 텍스트 합성 키프레임을 표시할 부분의 크기가 크면 상기 합성을 위해 선택되는 텍스트의 크기나 개수를 늘리고, 디바이스나 브라우저의 크기가 작고 텍스트 합성 키프레임을 표시할 부분의 크기가 작으면 상기 합성을 위해 선택되는 텍스트의 크기나 개수를 줄이는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 12】

제 1 항에 있어서, 상기 합성 텍스트 키프레임이 하나의 뉴스 기사 구간을 대표하는 키프레임으로서 생성되는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 13】

제 12 항에 있어서, 상기 합성 텍스트 키프레임이 뉴스를 앵커 샷 단위로 하나의 기사로 구분하여 그 뉴스 기사 구간을 대표하는 키프레임으로서 생성되는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 14】

뉴스 비디오에서 검출된 텍스트가 앵커 샷 텍스트인지, 에피소드 샷 텍스트인지를 판별하는 단계, 상기 뉴스 비디오에서 검출된 텍스트가 앵커 샷 텍스트인 경우를 에피소드 샷 텍스트인 경우 보다 더 높은 중요도 가중치를 할당하여 각각의 텍스트 중요도를 구하는 단계, 상기 구해진 텍스트 중요도가 일정 임계치 이상인 텍스트를 선택하고 이 선택된 텍스트를 하나의 이미지에 합성하여 하나의 뉴스 기사 구간을 대표하는 텍스트 합성 키프레임으로 생성하는 단계를 포함하여 이루어지는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 15】

제 14 항에 있어서, 상기 에피소드 샷 텍스트는 에피소드 샷 텍스트의 크기와 디스플레이 지속시간, 에피소드 샷 텍스트 내에서의 텍스트 요소의 평균 크기

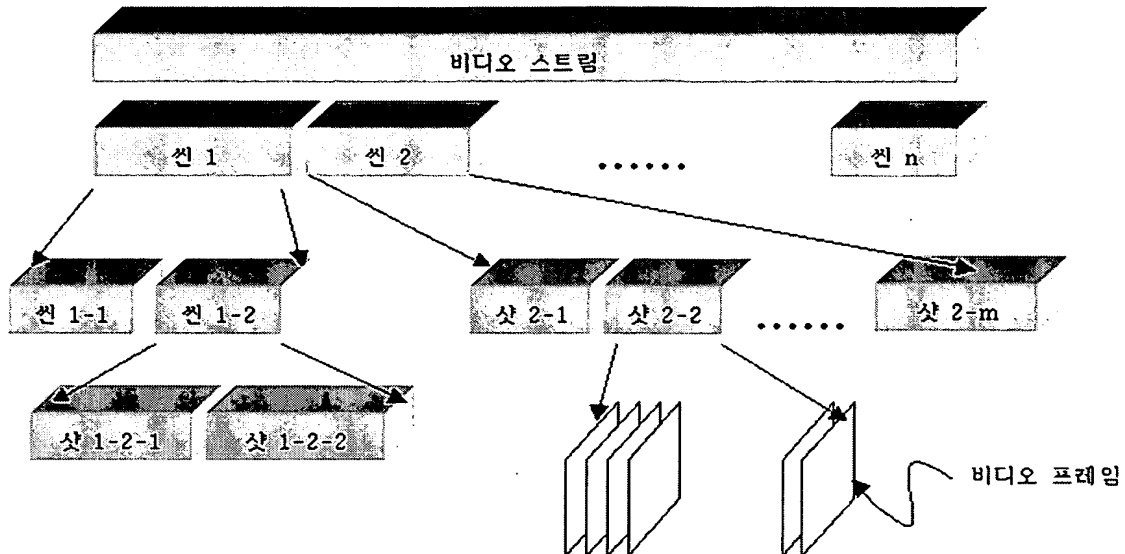
에 비례하여 중요도를 높이는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【청구항 16】

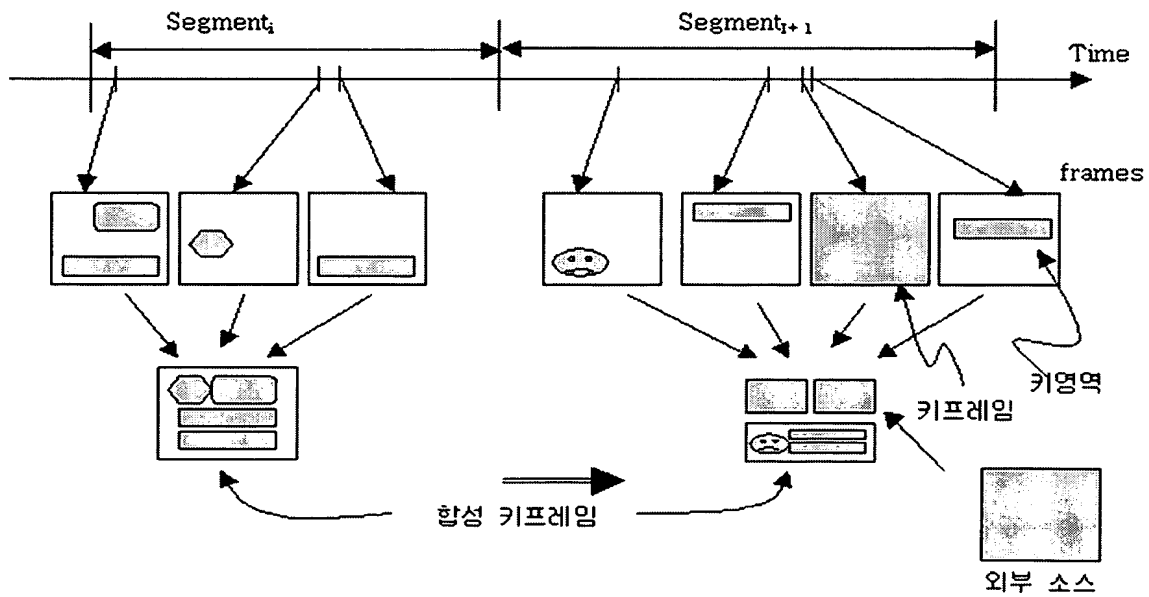
제 15 항에 있어서, 상기 에피소드 샷 텍스트의 중요도는 그 텍스트의 크기와 디스플레이 지속시간, 텍스트 요소의 평균 크기에 대하여 각각 다른 가중치를 할당하여 해당 에피소드 샷 텍스트에 대한 중요도를 구하는 것을 특징으로 하는 비디오 텍스트 합성 키프레임 생성방법.

【도면】

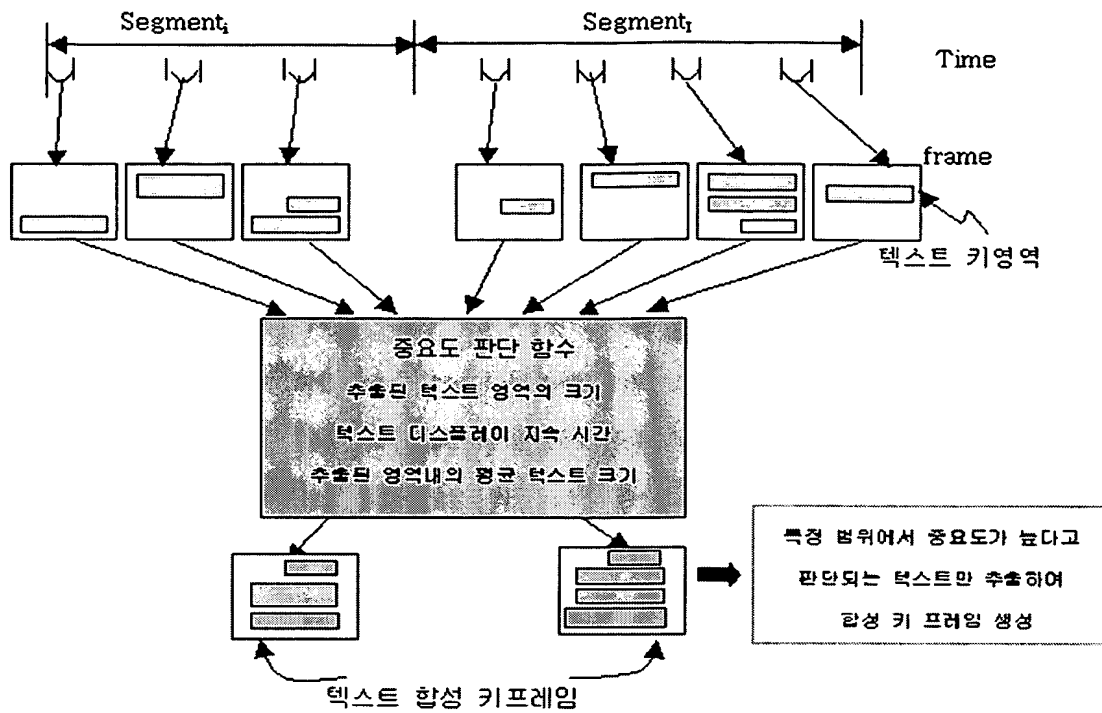
【도 1】



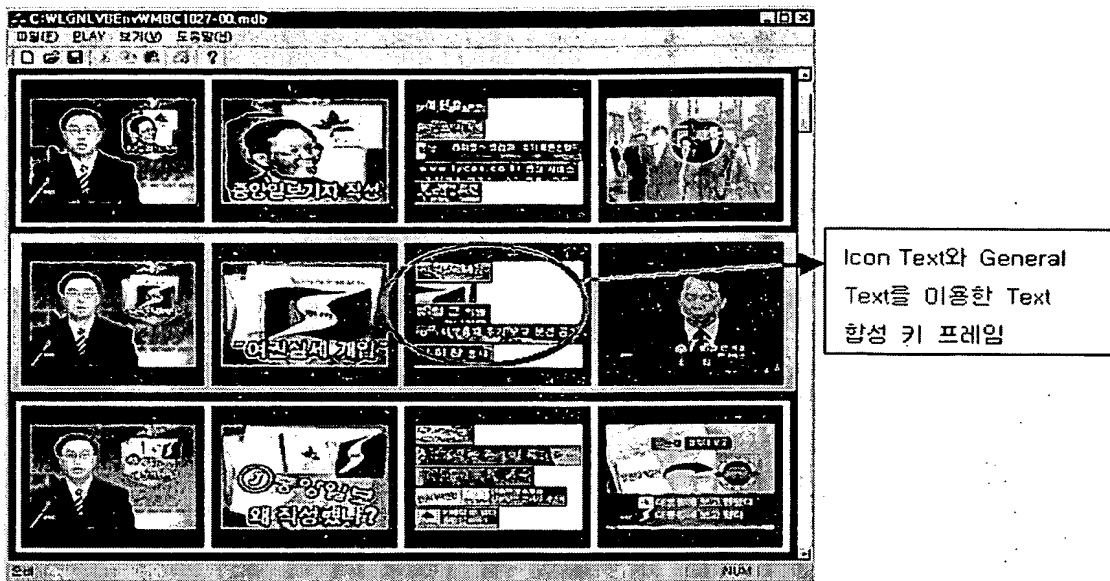
【도 2】



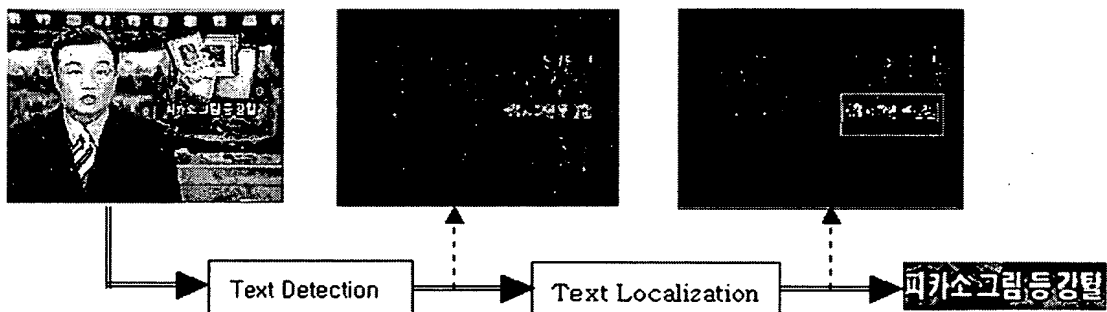
【도 3】



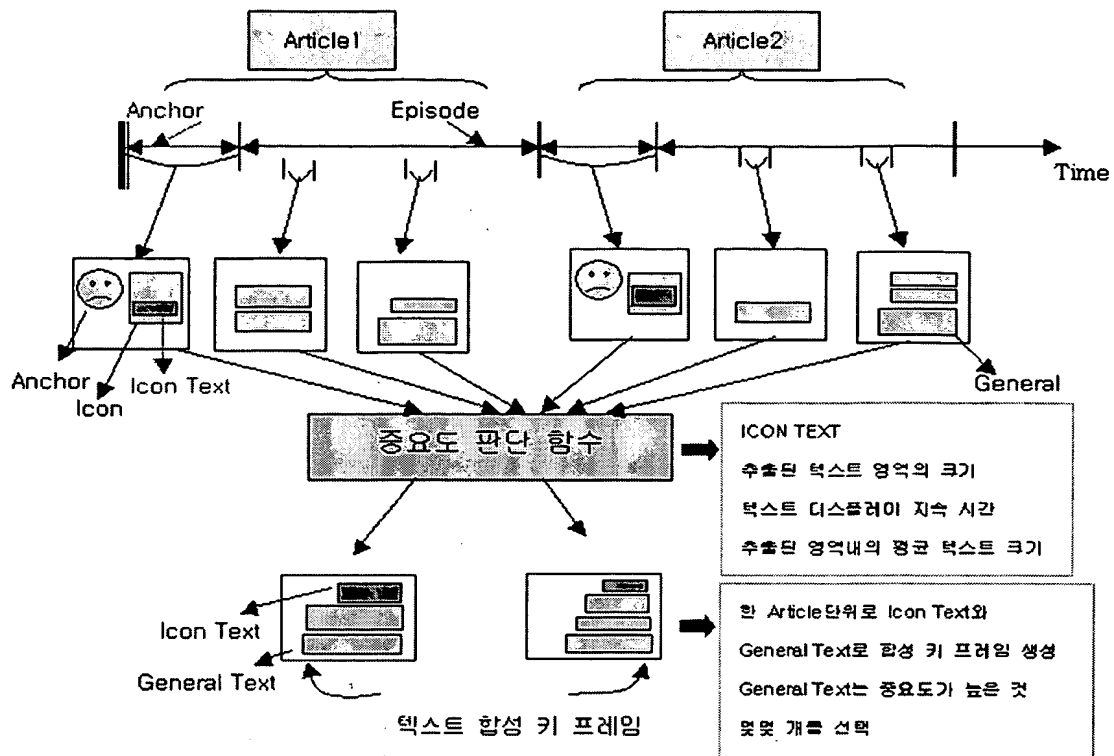
【도 4】



【도 5】



【도 6】



【도 7】

